

ノードとスライダで細部調整を追い込む画像生成システム

大友 千宙* 宮下 芳明*

概要. Text-to-Image モデルでの画像生成では、ユーザはテキストのみで生成したい画像を指示する。生成された画像の中には、細部の編集により表現を追い込みたいものがある。だが、これまでに提案されてきた画像生成のための編集手法では、表現を追い込む際に必要なインタフェースはあまり提供されていない。また、ノード内のパラメータを調整することで画像を生成するビジュアルプログラミングのようなシステムが多く開発されているが、生成した画像を編集できるものは少ない。本稿の目的は画像の生成と編集が一体化したシステムによって、ユーザ自らが求める表現を追求できるようにすることである。そこで、本稿ではテキストボックスとスライダからなるノードの組み合わせでプロンプトを構築し、テキスト変更やスライダ調整によって画像の生成と編集を行う画像生成システムを提案する。ユーザはテキストの操作で画像内オブジェクトの部分的な入れ替え、スライダの操作で特定のオブジェクトの影響力を微調整するといった「詰める」作業が可能となる。提案システムでの「詰める」作業によって、ユーザが求める表現に到達できる。

1 はじめに

Stable Diffusion[7]などで生成した画像の中には、自分が望む画像になるように細部を編集して追い込みたいものがある。例として、ユーザの想定よりも白みがかかった「夕暮れの風景画」が生成された場合を考える。このとき、プロンプト内の「夕暮れ」を操作しより赤みがかかった「夕暮れの風景画」が生成できれば、細部を追い込むことでユーザが望んでいた表現を追求できたといえる。

だが、既存の画像生成向けの編集手法はプログラミングにより利用する必要がある、表現を追い込むためのインタフェースを提供しているものは少ない。また、ノード内のパラメータを細かく調整することで画像を生成するビジュアルプログラミングのようなシステムが提案されているが、生成した画像を編集できるものは少ない。

表現を追い込む作業では、値を細かく調整できるスライダ等のインタフェースが必要となる。だが、現状では追い込むためのインタフェースと編集手法を兼ね備えた画像生成システムが少ない。そのため、表現を「詰めたい」ユーザが一つのシステム内で画像の生成と編集を行うことは困難である。

本稿では画像生成においてユーザ自身が求める表現を追求できるようにすることを目的とし、テキストボックスとスライダを有したノードでプロンプトを構築し、個々のノードを調整することで画像の生成と編集が行えるシステムを提案する。ユーザはノードの操作によって元画像の構図や内容を維持したまま画像内オブジェクトの部分的な入れ替え、特定の

オブジェクトの影響力を微調整するといった「詰める」作業が可能となる。表現を追い込むためのインタフェースによって画像の生成と編集が可能となるため、ユーザが生成した画像を「詰める」ことで求めていた表現に到達することが期待できる。

2 関連研究

Text-to-Image モデルによって生成される画像をテキスト操作のみで制御するための手法が研究されている。Hertzらはテキスト操作のみで生成する画像の構図と内容を制御する手法、prompt-to-promptを提案した[5]。また、Tumanyanらは入力した画像の構図を保ったままユーザが入力したテキストの内容に従う画像を生成できる手法、Plug-and-Playを提案した[8]。この2つの手法を基に、Geらはリッチテキストによって生成される画像の構図や内容を制御する手法を提案した[4]。どの手法もテキストによる画像生成の制御を目的としており、表現を追い込むためのインタフェースを提供するものではない。

また、ビジュアルプログラミングと画像生成を組み合わせたシステムが多く提案されている。ComfyUIはノード接続によってStable Diffusionのワークフローを構築するシステムである[2]。ComfyUIは処理を細かく制御したい人を対象としているため、前述した画像生成の制御手法には対応していない。VPGenは大規模言語モデルによってテキストからオブジェクトの情報を取得し、画像を生成するフレームワークである[1]。テキストのみで生成する画像の構図を細かく制御できるが、画像の編集は出来ない。Ultraforgeは画像生成と編集のためのノードベースツールである[3]。生成した画像にぼかしなどの処理をかけることが可能だが、ComfyUIと同様に生成された画像を編集する手法に対応していない。

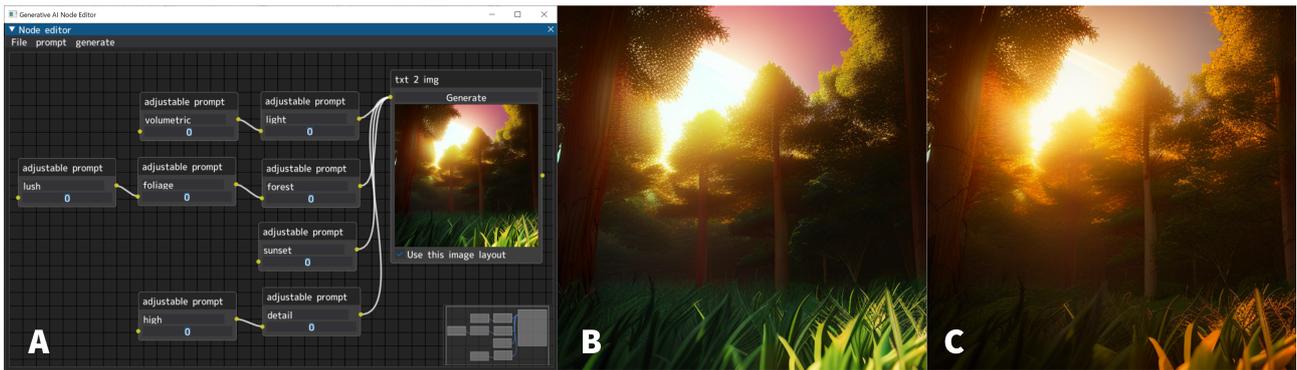


図 1: (A) 提案システムのスクリーンショット
 (B) 提案システムによる編集を行う前の画像
 (C) 提案システムによって細部編集を追い込んだ画像

3 提案システム

提案システムはノードの操作を行うノードエディタと、プロンプト作成や画像の生成と編集を行う Generative AI からなる。提案システムのスクリーンショットを図 1A に示す。実装には Python のライブラリである DearpyGUI[6] と Python3.10 によって実装した。加えて、DearpyGUI によって作成された画像処理ツールである Image-Processing-Node-Editor[9] のソースコードを一部利用した。画像の生成はプロンプトを構築するためのプロンプトノードを画像を生成するための生成ノードに接続することで行う。生成した画像の編集はノードの追加、テキストの入れ替え、スライダの調整によって行う。以下では、この 3 つの操作によってどのように画像を編集できるかについて述べる。

ノードの追加により、画像内にあるオブジェクトの属性を指定したり、全体の画風を指定したりすることが可能である。具体的に、「white」と入力したノードを「hat」と入力済みのプロンプトノードに接続して生成したとする。すると、生成された画像の背景などはそのままに、帽子のみが白くなる。このように、既存のプロンプトノードに新たなノードを接続することで既存ノードの属性を指定することができる。また、生成ノードに直接プロンプトノードを追加することで画像全体の画風を指定できる。

テキストの入れ替えにより、生成した画像の構図を維持したまま画像内のオブジェクトを入れ替えることができる。例えば、既存プロンプトノードの「cat」を「dog」に書き換えて生成すると、すでに生成されていた画像内の猫が犬に入れ替えることができる。この時、帽子や背景といったオブジェクトは大きく変化しないため、画像内の特定部分だけを入れ替えることが可能である。

プロンプトノード内のスライダを調整することで、そのプロンプトノードが持つテキストが生成結

果に与える影響を変更することができる。スライダは左方向と右方向にそれぞれ 10 段階動かすことが可能である。ノードを追加した段階のスライダの値は 0 となっている。一例として、テキストボックスに「night」と入力されているノードのスライダを右方向に操作したとする。すると、画像内における「night」の影響が増すため画像全体の明るさが減少する。また、スライダを左方向に動かすとそのノードが生成結果に与える影響を減少させることができる。例えば、「night」ノードのスライダを左方向に動かすと、画像全体の明るさが増加する。

4 提案システムによる作例

提案システムを用いて生成した画像と、その画像の細部を編集して追い込んだ作例を図 1B と図 1C に示す。また、図 1A は図 1B を生成した時の提案システムのスクリーンショットである。これらの作例は、ユーザが「夕暮れの森」を題材として画像を制作する場合を想定した。そのため、図 1A からは「sunset」や「forest」などの単語がノードに入力されていることが分かる。

ここで、図 1B はユーザが想定していた「夕焼けの森」よりも「夕焼け」があまり画像に反映されていないと感じたと仮定する。こうした時、提案システムではあまり反映されていないと感じる箇所に該当するノードのスライダを操作することで、画像全体の構図や内容を保ったまま該当箇所をより強く画像に反映することができる。図 1C は実際に図 1A 内の「sunset」と入力されたノードのスライダを右方向に動かして制作したものである。

図 1B は白みがかかった光によって全体が照らされているのに対し、図 1C は赤みがかかった光で全体が照らされている。よって、ユーザは提案システムによって特定の部分を「詰める」ことでより求めている表現を追求した画像を得ることができるといえる。

参考文献

- [1] J. Cho, A. Zala, and M. Bansal. Visual Programming for Text-to-Image Generation and Evaluation. In *NeurIPS*, 2023.
- [2] comfyanonymous. ComfyUI, March 2023. <https://github.com/comfyanonymous/ComfyUI>.
- [3] F. Forge. Announcing Ultraforge, June 2023. <https://www.filterforge.com/ultraforge/>.
- [4] S. Ge, T. Park, J.-Y. Zhu, and J.-B. Huang. Expressive Text-to-Image Generation with Rich Text. In *IEEE International Conference on Computer Vision (ICCV)*, 2023.
- [5] A. Hertz, R. Mokady, J. Tenenbaum, K. Aberman, Y. Pritch, and D. Cohen-Or. Prompt-to-Prompt Image Editing with Cross Attention Control. *arXiv preprint arXiv:2208.01626*, 2022.
- [6] J. Hoffstadt, P. Cothren, et al. DearPyGUI.
- [7] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, June 2022.
- [8] N. Tumanyan, M. Geyer, S. Bagon, and T. Dekel. Plug-and-play diffusion features for text-driven image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1921–1930, 2023.
- [9] 高橋かずひと. Image-Processing-Node-Editor, June 2022. <https://github.com/Kazuhito00/Image-Processing-Node-Editor>.
- [10] 大友千宙, 宮下芳明. 研究概要に基づいて質問を生成する AI システム. WISS 第 30 回インタラクティブシステムとソフトウェアに関するワークショップ, pp. 1–3, 2022.

未来ビジョン

WISS2022において、著者らは論文の概要を入力すると GPT-3 が生成した質問を出力するシステムを提案した [10]。この研究では、自身が所属する研究グループ内からは得られない質問を発表本番で得るためには、発表準備の段階から客観的な想定質問が必要だと考えた。そこで、LLM を第 3 者として捉え、生成された質問に対する回答を発表内に含めることで自身の発表を改善することを促した。この研究は、Generative AI とのやり取りを経て自身の研究発表という表現をより追い込むことを提案した研究ともいえる。

本稿では Generative AI を表現のための道具として捉え、表現を追い込む際に有効なインタフェースにより操作するシステムを提案した。現時点における Generative AI の主なインタフェースはテキストである。インタフェースとしてのテキストは、ありとあらゆるパラメータの数値を一つずつ変えながら表現を追い込むといった作業には不向きである。表現を追い込むためのシステ

ムは、Unity のシェーダーグラフや DaVinci Resolve の color correction などに代表されるノードベースのシステムが多い。このようなシステムはノード内部にあるスライダなどを操作することで表現を「詰める」ことができる。提案システムをスライダを有するノードによるノードベースシステムとしたのは、画像生成という表現のための道具によって画像という表現を「詰める」ことを目的としたためである。

また、プロンプトをノードという単位に分割し、後から編集するためのインタフェースを設けるという枠組みは、画像以外のコンテンツ生成にも応用できると考えている。なぜなら、現時点における Generative AI の主なインタフェースはテキストであるためである。本稿の提案システムが、Generative AI を表現のための道具として扱う際に表現を「詰める」際に有効な 1 手法となることを期待している。